Woot, an Active Gypsy-Class Retrotransposon in the Flour Beetle, *Tribolium castaneum*, is Associated With a Recent Mutation

Richard W. Beeman,* M. Scott Thomson,[†,1] John M. Clark,[†,2] Marco A. DeCamillis, Susan J. Brown[†] and Robin E. Denell[†]

**U.S. Grain Marketing Research Laboratory, U.S. Department of Agriculture, ARS, Manhattan, Kansas 66502 and* [†]*Division of Biology, Kansas State University, Manhattan, Kansas 66506*

Manuscript received November 27, 1995
Accepted for publication January 26, 1996

ABSTRACT

A recently isolated, lethal mutation of the homeotic *Abdominal* gene of the red flour beetle *Tribolium castaneum* is associated with an insertion of a novel retrotransposon into an intron. Sequence analysis indicates that this retrotransposon, named *Woot*, is a member of the *gypsy* family of mobile elements. Most strains of *T. castaneum* appear to harbor ~25–35 copies of *Woot* per genome. *Woot* is composed of long terminal repeats of unprecedented length (3.6 kb each), flanking an internal coding region 5.0 kb in length. For most copies of *Woot*, the internal region includes two open reading frames (ORFs) that correspond to the *gag* and *pol* genes of previously described retrotransposons and retroviruses. The copy of *Woot* inserted into *Abdominal* bears an apparent single frameshift mutation that separates the normal second ORF into two. *Woot* does not appear to generate infectious virions by the criterion that no *envelop* gene is discernible. The association of *Woot* with a recent mutation suggests that this retroelement is currently transpositionally active in at least some strains.

THE use of modified transposable elements for such techniques as germ line transformation, cloning by transposon tagging, and enhancer detection have contributed dramatically to the power of *Drosophila melanogaster* as a genetic model system. Such genetic tools would be equally valuable in other insect model systems and for the manipulation and possible control of insect pests and vectors of disease. So far, attempts to utilize Drosophila elements in other insect taxa have been unsuccessful (O'BROCHTA *et al.* 1994). Furthermore, attempts to identify technically useful endogenous elements have been unsuccessful to date, in part because purely molecular approaches do not discriminate between inert and functional copies (*e.g.*, ROBERTSON 1993). The history of research on Drosophila suggests that an alternative strategy will be more successful: only those elements that are currently transpositionally active will be ascertained by organismal genetic approaches, such as detection of spontaneous insertional mutations or other evidence of hybrid dysgenesis.

The red flour beetle, *Tribolium castaneum*, offers the most facile insect system for genetic analysis currently available outside of the Drosophilids (*e.g.*, BEEMAN *et al.* 1992a; STUART *et al.* 1993). We have been taking two approaches to find mobile transposable elements in this species. One approach has been to examine crosses between beetle strains collected from all over the world for evidence of hybrid dysgenesis, such as sterility or semisterility. This tack has resulted in the description of a novel class of maternally acting selfish genetic elements (BEEMAN *et al.* 1992b), as well as an apparent transposable element with mobility restricted to the soma (THOMSON *et al.* 1995), but it has not yet led to the identification of a potentially useful element mobile in the germ line. A second approach involves examining recently isolated spontaneous mutations for the insertion of transposable elements. We show here that a variant of the homeotic gene *Abdominal* is associated with the insertion of a retrotransposon with interesting properties. It is likely that the further implementation of this strategy will identify additional and potentially useful elements.

Corresponding author: Richard W. Beeman, U.S. Grain Marketing Research Laboratory, USDA, ARS, 1515 College Ave., Manhattan, KS 66502. E-mail: beeman@crunch.usgmrl.ksu.edu

[1] *Present address:* Department of Biological Sciences, University of Wisconsin-Parkside, Kenosha, WI 53141.

[2] *Present address:* Department of Entomology, University of Massachusetts, Amherst, MA 01003.

MATERIALS AND METHODS

Tribolium strains and culture: GA-1 is a standard wild-type strain maintained in our laboratory since 1980 (HALISCAK and BEEMAN 1983). A^4 is a spontaneous null mutation in the homeotic *Abdominal* gene (BEEMAN *et al.* 1989; STUART *et al.* 1993). The A^4 mutation was found in the GA-1 stock (see below) and was established as a balanced lethal stock using the balancer chromosome, *Eyeless* (*Ey*) (BEEMAN *et al.* 1995). We used the wild-type strain of *T. freemani* described by NAKAKITA *et al.* (1981). Wild-type *T. brevicornis* were obtained from cultures maintained for many years at the Manhattan laboratory. Origins of other strains are given in the figure legends. Strains were maintained at 30° on whole wheat flour fortified with 5% brewers' yeast.

Phage libraries and subcloning: A genomic library was con-

structed in λGEM-11 (Promega) using DNA from the A^4/Ey strain of *T. castaneum* after partial digestion with *Sau*3A as described (BROWN *et al.* 1990). Following the supplier's directions, genomic fragments were end-filled and ligated into λGEM-11 *Xho*I half-site arms. The library was screened initially with *Abdominal* (*A*) probes previously cloned from a GA-1 genomic library, and the 12-kb A^4 insertion was cloned by subsequent chromosome walking. *Woot* restriction fragments were subcloned into pGEM7F+ (Promega) or M13mp18/19 for further analysis. The rearrangement associated with the A^4 mutation was localized by comparison of the restriction maps from GA-1 and A^4 chromosomes. A GA-1 embryonic cDNA library (STUART *et al.* 1993) was screened with a ^{32}P-labeled mixture of subcloned fragments encompassing the 5′ long terminal repeat (LTR) and central body of *Woot*.

DNA sequence determination and analysis: For subcloned inserts, a series of nested deletions was constructed (HENIKOFF 1984) and sequenced by the dideoxynucleotide chain-termination method (SANGER *et al.* 1977). The entire *Woot* 5′ LTR and the central body were sequenced on both strands, and the 3′ LTR was sequenced on at least one strand. DNA and protein sequence compilation and analysis were done with the aid of the AssemblyLign/MacVector software package (IBI, Inc.). Online searches of nucleic acid and protein sequence databases were conducted using the BLAST algorithm (ALTSCHUL *et al.* 1990).

PCR cloning: Primers (forward 5′-AAAATGGAGATGGGGACG-3′ and reverse 5′-ACCCGTTGAGTTCTCGCATC-3′) were used to amplify by PCR a 656-bp fragment from genomic DNA of *T. castaneum*, *T. freemani* and *T. brevicornis*. The amplified region lies near the boundary of the Ribonuclease H and Integrase domains of *Woot*. Amplified fragments were separated on low-melting agarose gels, eluted using PCR Wizard prep columns (Promega), ligated into the TA cloning vector pCR3 (Invitrogen), cloned and subjected to cycle sequencing using the fmole kit (Promega).

PCR amplification of specific alleles (PASA): The occurrence of the open reading frame (ORF) 2 frameshift mutation in *Woot* copies was assessed by PASA. An allele-specific reverse primer was made by placing the frameshift mutation at the 3′ end of the reverse primer 5′-TGTCGGCAACTACATTCC-3′. The allele-independent forward primer was the same as that used for PCR cloning described above. These primers should amplify a 447-nucleotide fragment from *Woot* elements containing the frameshift mutation.

Southern hybridization: To examine the presence of *Woot* elements in various strains of *T. castaneum*, genomic DNA was isolated and digested to completion with *Eco*RI. The resulting blots were probed at high stringency with a gel-purified 4.4-kb *Hin*dIII fragment (probe B in Figure 1) as follows. Filters were prehybridized for 1 hr at 65° in 10× Denhardt's, 6× SSC, 0.1% SDS, 25 mM phosphate buffer pH 7.0 and 0.1 mg/ml denatured sonicated herring sperm DNA. The filters were hybridized for 24 hr under the same conditions, then washed twice at 65° with 2× SSC, 0.1% SDS. Probe B spans the Integrase domain and most of the 3′ LTR and is predicted to hybridize with fragments containing 3′ or 5′ junctions after *Eco*RI digestion.

RESULTS

The origin of A^4: The Tribolium homeotic gene *Abdominal* (*A*) is the homologue of *abdominal-A* in Drosophila (BEEMAN *et al.* 1989; STUART *et al.* 1993). Beetles homozygous for some hypomorphic alleles survive to adulthood and display homeotic transformations of the anterior abdomen. In homozygous condition, null mutations result in embryonic lethality and anteriorward

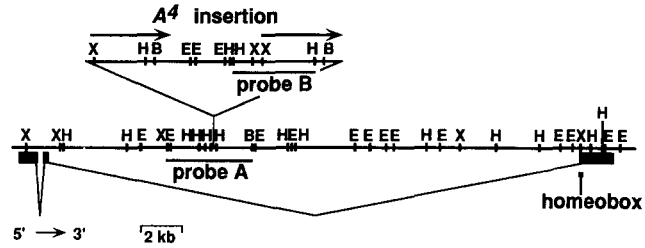

FIGURE 1.—Molecular structure of the *Abdominal* gene of wild-type strain GA-1 of *T. castaneum*, showing the *Woot* insertion point, as well as the restriction map of the *Woot* insertion. *Abdominal* exons are indicated by thick lines and introns by thin lines. Restriction endonuclease abbreviations: E, *Eco*RI; H, *Hin*dIII; B, *Bam*HI; X, *Xho*I. The insertion (~12 kb) is drawn to scale. The wild-type *Bam*HI-*Xho*I fragment (probe A) was used to isolate *Woot*-bearing fragments derived from A^4 mutant chromosomes. Arrows indicate the *Woot* LTRs and also the direction of transcription for both the *Woot* element and the *Abdominal* gene into which it is inserted. The *Woot* H fragment (probe B) was used for Southern hybridization analysis of *Woot* junction fragments (see Figure 7).

transformations of the entire abdomen. Individuals heterozygous for null mutations and the hypomorphic variant *Abdominal-missing abdominal sternites* (A^{mas}) also survive to yield adults with transformed abdomens. In an experiment designed to generate new *A* alleles, five groups of 40 gamma-irradiated (4 kR) virgin males from the GA-1 strain were each mass-crossed with 100 virgin females homozygous for A^{mas} and the closely linked homeotic variant *maxillopedia* (BEEMAN *et al.* 1989). Two groups yielded single new *A* variants recognized by the failure to complement A^{mas}, whereas a third group (designated C) generated an apparent cluster of 16 adults bearing new mutations. There were ~2200 F_1 progeny from each group or an average of ~55 progeny from each irradiated male. If one of the males in group C had been heterozygous for a preexisting *Abdominal* mutation, the expected number of affected F_1 progeny would have been 27, assuming an equal contribution from each male and no viability reduction associated with the mutation. Although these data do not preclude a premeiotic mutational event in an irradiated male, they are consistent with the spontaneous origin of a mutation in a previous generation. The dominant visible mutation *Eyeless* (*Ey*) is associated with crossover suppression that makes it a balancer for the *Abdominal* region. Thus, one of the mutant individuals from group C was crossed to an *Ey* mate, and a balanced stock was generated. The new variant was designated A^4. No reversion or other instability of the mutation has been observed during the ~30 generations that this stock has been maintained.

Characterization of the A^4 mutant lesion: Most of the *Abdominal* gene, including the entire coding region, has been molecularly characterized (STUART *et al.* 1993; S. J. BROWN, unpublished results), and its organization is shown in Figure 1. Genomic DNA from A^4/Ey beetles was systematically probed with subclones from this re-

gion in a search for a restriction fragment length polymorphism (RFLP) with respect to GA-1 DNA. Southern hybridization analysis of genomic DNA digested with either *Bam*HI, *Eco*RI, or *Hin*dIII and probed with a *Xho*I-*Bam*HI fragment (probe A, see Figure 1) revealed such an RFLP. As indicated, this probe is derived from a large intron ~15–20 kb upstream of the homeobox-containing exon. Analysis of restriction fragment data indicated that A^4 was associated with a chromosomal rearrangement within the region used as a probe.

An A^4/Ey genomic library was constructed and screened with probe A, and restriction mapping was used to identify those isolated clones that included the rearranged DNA. The results (Figure 1) showed that the A^4 mutation was associated with a ~12-kb insertion. Moreover, the restriction map indicated the presence of long terminal direct repeats (arrows in Figure 1), indicating that the insertion was probably a transposable genetic element. This element is hereafter referred to as *"Woot,"* a reference to the character *Woot* the Wanderer in Frank Baum's classic "The Tin Woodman of Oz."

Organization of the *Woot* element: Figure 2 compares the organization of the *Woot* element to that of the Drosophila retrovirus *gypsy* (MARLOR *et al.* 1986). In common with retroviruses and some retrotransposons (VARMUS and BROWN 1989), *gypsy* has LTRs of ~500 bp. Within each LTR, an R region separates the repeat into flanking U3 and U5 regions. Transcription is initiated at the beginning of the upstream R, and a polyadenylation site marks the end of the downstream R. The resulting transcript is translated or serves as the viral genome. *Gypsy* also resembles retroelements in general in having an ORF called *gag*, encoding a polyprotein contributing nucleocapsid components, and a second (termed *pol*) encoding a polyprotein supplying Protease, Reverse Transcriptase, RNase H and Integrase functions. In at least some well characterized cases, a ribosomal frameshift allows translation of the *pol* polyprotein (HATFIELD *et al.* 1992). A third ORF encodes an *envelope* protein. The presence of an *envelope* gene (characteristic of vertebrate retroviruses as well) is apparently important in the functioning of *gypsy* as an infectious virus (KIM *et al.* 1994; SONG *et al.* 1994; TANDA *et al.* 1994). In contrast, most otherwise closely related elements are retrotransposons lacking an *envelope* gene. *Gypsy* resembles other infectious retroviruses in that the transcript encoding *env* is generated by splicing out ORFs 1 and 2 (PELISSON *et al.* 1994). Also indicated in Figure 2 are tRNA and oligopurine primer binding sites (PBSs) necessary for reverse transcription and second strand synthesis.

The sequence of the A^4 *Woot* element (Figure 3) reveals many features that identify it as a retrotransposon inserted in the same transcriptional orientation as the *Abdominal* gene (see also Figure 1). The 5′ LTR and the central body have been completely sequenced (GenBank accession number U09586). Sequencing of the 3′ LTR on at least one strand suggests that it is

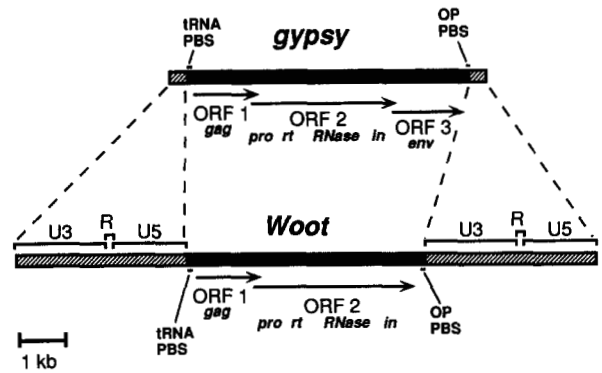

FIGURE 2.—Organization of the *gypsy* retrovirus compared to that of the *Woot* retrotransposon. Hatched segments represent LTRs; while solid segments represent retroelement bodies. Arrows delimit ORFs and indicate direction of transcription. The (rare) frameshift mutation that interrupts *Woot* ORF 2 in the A^4 copy is not shown (see text for explanation). OP, oligopurine; PBS, primer binding site for reverse transcription or second strand synthesis; gag, group-specific antigen; pro, protease; rt, reverse transcriptase; RNase, ribonuclease H; in, integrase; env, envelop; U3, unique 3′ region of transcript; R, redundant region; U5, unique 5′ region.

identical to the 5′ LTR. Primer binding sites for cDNA synthesis are present at the termini of the central body (Figures 2 and 3). Conceptual translation of the A^4 *Woot* element central body identifies three ORFs, designated as ORFs 1, 2a and 2b. The latter pair are combined as ORF 2 in Figure 2. These ORFs would encode proteins of 452, 712, and 470 amino acids, respectively. As described below, most *Woot* elements have a single ORF 2 that includes both 2a and 2b and encodes a protein of 1158 amino acids. The copy inserted into the *Abdominal* gene has an apparent frameshift mutation.

Portions of the predicted product of ORF 2 show the strongest conservation among retroelements (DOOLITTLE *et al.* 1989; XIONG and EICKBUSH 1990). Sequence comparisons identify a *gypsy* group, including *gypsy* and some other LTR-containing retrotransposons and putative retroviruses. Searches of protein databases using the *Woot* ORF 2 predicted protein as query revealed extensive similarity with other elements of the *gypsy* group. Figure 4 shows the conceptual translation of a 1080-nucleotide segment of the A^4 *Woot* ORF 2a (nt 5871–6950 in Figure 3) aligned with the corresponding regions of *pol* gene products from the *gypsy* group elements *gypsy, Ulysses,* and *mdg1* from Drosophila, *mag* from Lepidoptera, and *TY3* from yeast. Over the entire *pol* polyprotein, *Woot* most closely resembles the *D. virilis* retrotransposon *Ulysses*. As aligned in Figure 4, the two peptides are 33% identical and 55% similar or identical in amino acid sequence. In this region, the predicted *Woot* protein is much less similar to those encoded by cauliflower mosaic virus and Moloney murine leukemia virus, members of the two groups most closely related to the *gypsy* class (XIONG and EICKBUSH (1990). Other features of the predicted *Woot* ORF 2 polyprotein that are characteristic of retroelement Protease, Reverse

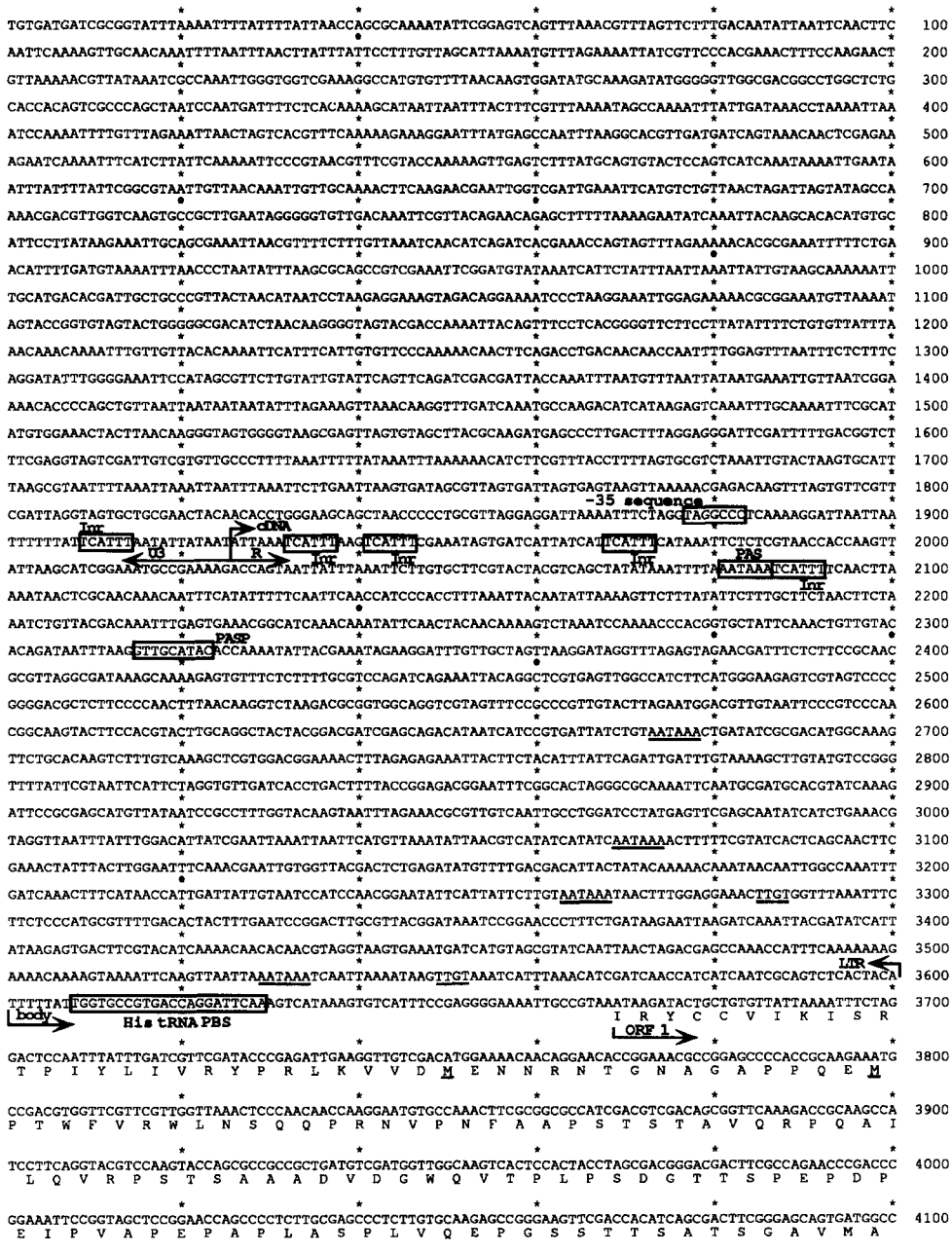

FIGURE 3.—Nucleotide sequence and conceptual translation of the *Woot* retrotransposon of *T. castaneum*. The 3′ LTR is not shown. Abbreviations: Inr, initiator sequence for transcription; PAS, polyadenylation signal; PASP, polyadenylation site potentiation motif; AP, aspartyl protease motif; RT, reverse transcriptase motif; IN, integrase motif. See Figure 2 legend for remaining abbreviations. The upstream PBS (nt 3608–3629) appears to utilize a histidine tRNA, since it is complementary (with only one mismatch) to the 3′-terminal 22 nucleotides of a his-tRNA of *D. melanogaster* (ALTWEGG and KUBLI 1980). Asparagine-rich region near C-terminus is indicated by underlines. Putative zinc finger motif (nt 7441–7557) is suggested by paired histidine and cysteine residues (underlined) at the predicted location within the integrase gene. The ORF 2 frameshift mutation (see Figure 5) is underlined at nt 7101.

Transcriptase, and Integrase proteins are indicated in Figure 3.

Two potential methionine initiation codons in the *Woot* ORF 1 are indicated in Figure 3. Presumably, the predicted translation product of *Woot* ORF 1 corresponds to the *gag* protein of typical retroelements. *Gypsy* class *gag* proteins have no conserved sequence motifs (ESCHALIER 1989). However, like other *gag* proteins (FRIESEN and NISSEN 1990), the predicted product of *Woot* ORF 1 is proline-rich (10%), especially in a region near the amino terminus. It also has an uneven distribution of asparagine residues, including a 34 amino acid near the C-terminus that is 50% asparagine. This region is part of a larger hydrophilic stretch that spans ~110 amino acid residues or about one-fourth of the ORF 1 predicted protein.

Insertions of transposable elements are invariably associated with direct duplications of a short target se-

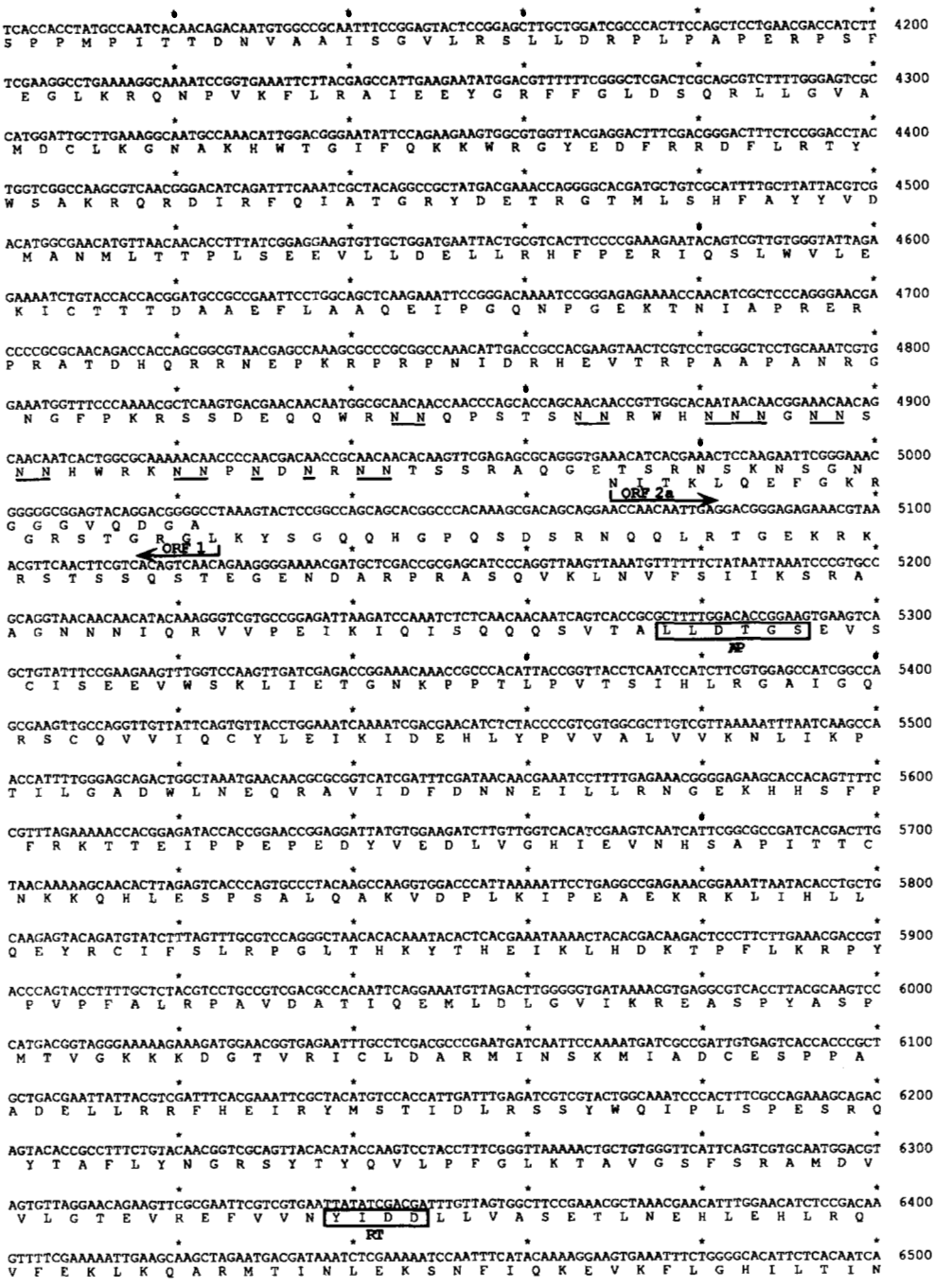

FIGURE 3.—*Continued*

quence adjacent to the point of insertion (VARMUS and BROWN 1989). The A^4 *Woot* insertion point is flanked by a repeat of the four-base sequence GTAC. Sequencing of the same region of a clone from a GA-1 genomic library revealed the sequence ATAC, indicating that the GA-1 stock is polymorphic in this respect. Such heterogeneity in a large intron is not unexpected, and polymorphisms for several other base pair differences and a single-base pair duplication/deletion were detected in the region immediately surrounding the insertion site. Typical of many retroelements (VARMUS and BROWN 1989), short complementary sequences (TCT at the 5′ end and ACA at the 3′ end) are found at the termini of the *Woot* transposon, directly adjacent to the target duplications.

Retroelement LTRs typically range in length from 300 to 500 bp (VARMUS and BROWN 1989). The 3600-bp *Woot* LTRs are the longest yet reported for any retroelement. The transcriptional organization of *Woot* has not been well characterized, but a cDNA has been isolated and sequenced. The organization of this cDNA is enigmatic. The 5′ terminus of this cDNA is in the LTR as expected if it were full-length or nearly so, but *Woot* sequence identity extends only to a position in ORF 2 between portions that encode the RNase H and Integrase functions. The remainder of the cDNA is

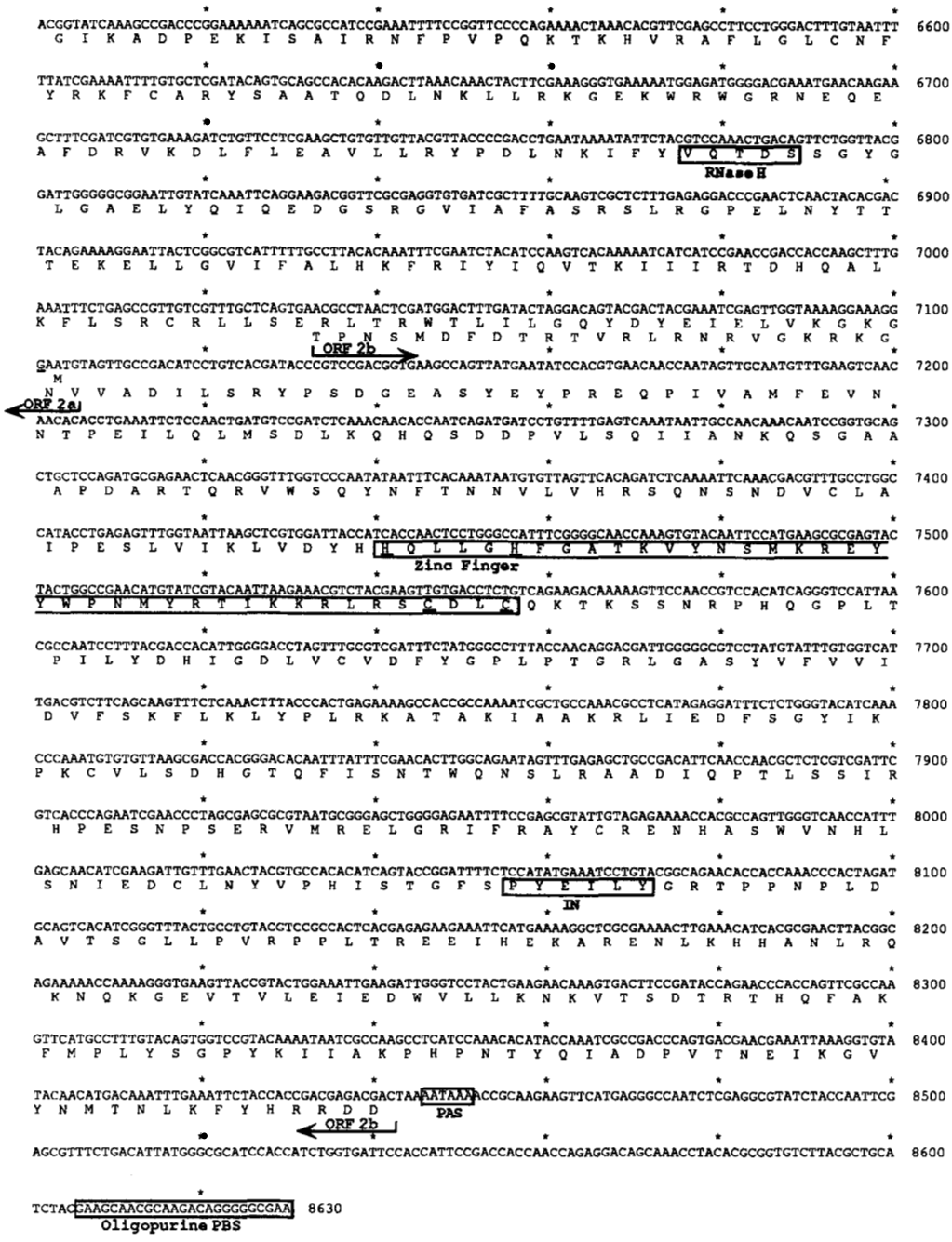

FIGURE 3.—*Continued*

comprised of ~30 nucleotides of divergent sequence and a polyA tract. There is a consensus polyadenylation signal 11 nucleotides upstream of the polyA tract, suggesting that the transcript that served as a template for first strand cDNA synthesis had undergone normal termination and 3′ processing. Possibly this transcript derived from a *Woot* element that had undergone a chromosomal rearrangement that juxtaposed the 3′ end of an unrelated transcription unit.

A basal promoter for Drosophila LTR retroelements includes an initiator element resembling TCAGTT at the RNA start site (ARKHIPOVA and ILYIN 1991; ARKHIPOVA 1995). The sequence TCATTT appears just upstream of the 5′ terminus of the *Woot* cDNA and is repeated twice within 20 nucleotides downstream, as well as twice more within 165 nucleotides downstream. Thirty-nine nucleotides upstream of the cDNA 5′ end is a potential "-35 sequence" (TAGGCCC), similar to the putative promoter sequences TACCCGG and TACGCGG found upstream of the transcription units of the *gypsy* class transposon *412* of *D. melanogaster* and the TAGGGG sequence of the Drosophila *gypsy* class transposon *mdg1* (YUKI *et al.* 1986). Although the spatial relationship of the 5′ terminus of this unusual cDNA with motifs of possible regulatory significance could be coincidental, it does

```
Woot   HDKTPFLKRPYPVPFALRPAVDATIQEMLDLGVIK-REASPYASPMTVGKKKDGTV------RICLDAR-   62
Ulys   EGAEPVKDRLNPLSPAKQEIVWAEVDKMLKLGIIE-ESDSPWSNRTTVV-MRPGKN------RFCLDAR-   61
TY3    GARLPRL-QPYHVTEKNEQEINKIVQKLLDNKFIV-PSKSPCSSPVVLVPKKDGTF------RLCVDYR-   61
MDG1   NDKTPVYIKNYRMPESQKPEIQRQVDKLIKDGIVE-QSISEYNSPLLLVPKKSLPNSEEKRWRLVVDYR-   68
Mag    PDAVPIYCRARPVPYALRERVDAELDAMLAAGVIKPVDHSDWATPLVVVRKADGGL------RICADYKV   64
Gyps   VDNEPVYSRAYPTLMGVSDFVNNEVKQLLKDGIIR-PSRSPYNSPTWVVDKKGTDAFGNPNKRLVIDFR-   68

Woot   MINSKMIADCESPPAADELLRRFHEIRYMSTIDLRSSYWQIPLSPESRQYTAFLYNGRSYTYQVLPFGLK  132
Ulys   KLNSVTVKDAYPLPCIEGILSRSTRLILSLASTLSSRSGNRDGGEEQGVYGVYCTRRPLYQFRHMPFGLC  131
TY3    TLNKATISDPFPLPRIDNLLSRIGNAQIFTTLDLHSGYHQIPMEPKDRYKTAFVTPSGKYEYTVMPFGLV  131
MDG1   QINKKLLADKFPLPRIEDILDQLGRAKYFSCLDLMSGFHQIELDERSRNITSFSTSTGAYRYTRLPFGLK  138
Mag    TLNKVLAIDRFPVPKMEDLFSNLSGNKFFTKLDLSQAYNQIVLSERSSEYTVINTHRGLFKYSRLVYGLA  134
Gyps   KLNEKTIPDRYPMPSIPMILANLGKAKFFTTLDLKSGYHQIYLAEHDREKTSFSVNGGKYEFCRLPFGLR  138

Woot   TAVGSFSRAMDVVLGTEVREFVVNYIDDLLVASETLNEHLEHLRQVFEKLKQARMTINLEKSNFIQKEV  201
Ulys   NAAQHFE-AHDKVIPANLRSNVFVYLDDLLIISADFPTHLKYLELVAECLKLRNANLTIGMAKSKFLFRNL  199
TY3    NAPSTFARYMADTF-RDLR-FVNVYLDDILIFSESPEEHWKHLDTVLERLKNENLIVKKKKCKFASEET  198
MDG1   IAPNSFQRMMTLAFSGLTPSQAFLYMDDLVVIGCSEKHMLKNLTDVF-KLCRQHNLKLHPEKCTFFMKEV  207
Mag    SSPGIFQKL-MVNMFKNVPNVVVFY-DDILIRNQDLDSHLKSIKEVLDILKERYGLKIKRSKCEFMVTEV  201
Gyps   NASSIFQRALDDVLREQIGKICYVYVDDVIIFSENESDHVRHIDTVLKCLKIDANMRVSQEKTRFFKESV  207

Woot   KFLGHI-LTINGIKADPEKISAIRNFPVPQKTKHVRAFLGLCNFYRKFCARYSAATQDLNKLL-------  263
Ulys   NYLGFIQLRRRTWRMDPGRVEAIRNIPNPRTVKELRSFLGTAGWYRRFIKNFAEISVPLTDAL-------  262
TY3    EFLGYS-IGIQKIAPLQHKCAAIRDFPTPKTVKQAQRFLGMINYYRRFIPNCSKIAQPI-QLF-------  259
MDG1   TYLGHK-CTDKGILPDDSKYEVIKNYPKPVNADEARRFVAFCNYYRRFIKNFSEKSRHLTRLC-------  269
Mag    RYLGFI-IDQNGVRVDPEKVKSIATMPHPNNVTELKSFIGMVNFYSKFIQDLSAHLSPLYALL-------  263
Gyps   EYLGFI-VSKDGTKSDPEKVKAIQEYPEPDCVYKVRSFLGLASYYRVFIKDFAAIARPITDILKGENGSV  276

Woot   -RK-GEKW--RWGRNEQEAFDRVK-DL-F-LEAVLLRYPDLNKIFYVQTDSSGYGLGAELYQIQEDGS-R  325
Ulys   -KKRTGRF--VLSDEAIEAIESLK-LA-L-TTAPVLVHADFRRPFFIQCDASHYGVGAVLFQLDDEQQ-E  325
TY3    --I-CDKS--QWTEKQDKAIEKLKAAL-C-NSPVLVPFNN-KANYRLTTDASKDGIGAVLEEVDNKNKLV  321
MDG1   -KK-NVPF--EWTSECNDVFEYLK-RK-L-MKPTLLQYPDFSKQFCITTDASKQACGAVLSQ-DHNGQQ-  330
Mag    -KK-GKHW--MWGNEQNAAFLNVK-KFLC-STKALAH-FDMSLESVLTVDASARGLGAVLAQRGPGCQ-E  325
Gyps   SKHMSKKIPVEFNETQRNAFQRLR-NI-LASEDVILKYPDFKKPFDLTTDASASGIGAVLSQ--E-G---  338

Woot   GVIAFASRSLRGPELNYTTTEKELLGVIFALHKFR  360
Ulys   RPIAFFSAKLNKHQINYSVTEKECLAAKLAIHRFR  360
TY3    GVVGYFSKSLESAQKNYPAGELELLGIIKALHHFR  356
MDG1   LPVAYASRSFTKGESNKSTTEQELAAIHWAINHFR  365
Mag    RVVAYASRALTTHELHYSQIHKEALAIVFAVEKFH  360
Gyps   RPITMISRTLKQPEQNYATNERELLAIVWALGKLQ  373
```

FIGURE 4.—Conceptual translation of a 1080-nucleotide segment of the A^4*Woot* ORF 2a (nt 5870–6949 in Figure 3) aligned with the corresponding regions of *pol* gene products from the *gypsy* group elements *Ulysses, mdg1,* and *gypsy* from Drosophila, *mag* from Lepidoptera, and *TY3* from yeast. Amino acid residues identical in at least five of the six elements are indicated in bold type. Dashes indicate gaps introduced to preserve the alignment. Genbank locus or accession numbers are as follows: *Ulysses,* DVULYSS; *mag,* BMMAG; *gypsy,* DROGYPF1A; *TY3,* S49799; *mdg1,* DMRTMGD1.

provide the basis for provisional assignment of the U3-R boundary.

The first potential polyadenylation signal AATAAA is found 155 nucleotides downstream of the putative transcription start site (Figure 3). Thus, the region ~15 nucleotides downstream of this signal is a potential polyadenylation site in the corresponding 3′ LTR, possibly defining the R-U5 boundary. Of the four other possible polyadenylation signals farther downstream, each of the last two are situated in appropriate proximity (~15 nucleotides upstream) to a potential transcription termination signal motif TTGT (underlined at positions 3285 and 3551 in the LTR, Figure 3). Determination of the exact position of the R-U5 boundary will require additional studies.

The A^4*Woot* has a frameshift mutation: ORFs 2a and 2b of the A^4*Woot* correspond to the 5′ and 3′ regions, respectively, of the *pol* gene of a typical *gypsy*-class retrotransposon. For various beetle strains, a portion of the *Woot pol* gene was amplified by PCR, cloned, and sequenced. Figure 5 shows an alignment of the A^4*Woot* sequence with those of two others independently derived from the balanced A^4 strain, as well as two independently amplified from the Tiw-1 strain, two derived from *T. freemani,* and one from *T. brevicornis.* At a position between the RNase H and Integrase coding regions, the A^4*Woot* copy has a single G in place of an AT common to each of the other copies (also see Figure 3). The presence of an AT in this position maintains an ORF, suggesting that the A^4*Woot* copy carries a frameshift mutation compared to other members of the *Woot* family representing three Tribolium species. The presence of this frameshift mutation in the A^4 strain but not in other strains was confirmed by allele-specific PCR. This approach utilized one primer that incorporated the A^4*Woot* frameshift at its 3′ end. Figure 6 shows that the expected fragment is amplified from the A^4 stock, but not from other strains of *T. castaneum* tested. The frameshift mutation was also absent from GA-1, the parent stock in which the A^4 mutation arose. The presence of this frameshift in the A^4*Woot* copy indicates that it cannot transpose autonomously, and that the Integrase function must be supplied in trans for transposition.

```
          G  K  G  -  M  *
             K  G  -N  V  V  A  D  I  L  S
Woot   GGAAAGGG-AATGTAGTTGCCGACATCCTGTCAC
G1     .......AT.....................
A1     .......AT.....................
A2     .......AT.....................
T1     .......AT.....................
T2     .......AT.....................
fre1   .......AT.....................G.
fre2   .......AT.....................G.
brv1   .......AT.....T.C.....T.....T....
brv2   .......AT.....T.C.....T.....T....
          G  K  D  N  V  V  A  D  I  L  S
```

FIGURE 5.—Sequences of *Woot* PCR fragments in the vicinity of the A^4 frameshift mutation. *Woot,* the *Woot* element derived from the A^4 mutant chromosome. G, A, T, fre and brv refer to the GA-1, A^4/*Ey* and Tiw-1 strains of *T. castaneum,* and to *T. freemani* and *T. brevicornis,* respectively. The Tiw-1 strain is identical to strain M in Figure 7. The numbers following the strain abbreviations indicate independent PCR clones. Dots indicate identity with *Woot.* The frameshift mutation in *Woot* is AT to G at nucleotide 8. The *Woot* translation in the two frames is shown above. Italics indicate amino acid residues derived by translation in the wrong frame (*, stop codon). The consensus translation is given below. The PCR temperature program was 30 cycles of 94° for 1 min, 50° for 2 min, 72° for 1 min 30 sec. See Figure 8 legend for primer sequences.

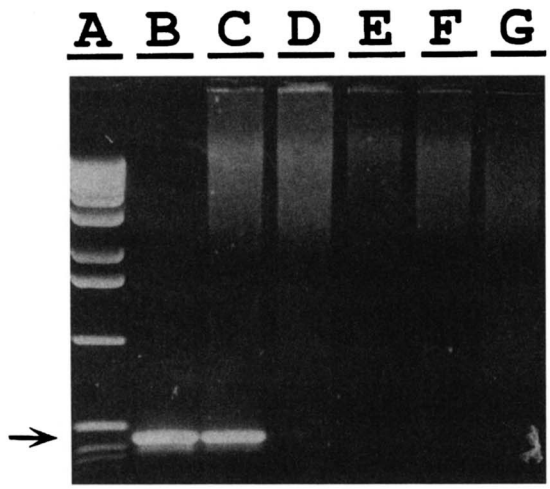

FIGURE 6.—Specific PCR amplification (PASA) of the ORF 2 frameshift mutation. Lane A, 1 kb ladder. Source strains for PCR template DNA and countries of origin: lane B, *Eco*RI/*Xho*I fragment of λJC05 subcloned into M13 (positive control); C, A^4/*Ey* (USA); D, Tiw-1 (India); E, *ab* (Colombia); F, Mek-1 (Peoples' Republic of China); G, GA-1 (USA). PCR conditions were as follows: 500 ng template (except lane B = 20 ng), 1 mM $MgCl_2$, 25 pmol each primer and 0.2 mM dNTPs in a final volume of 25 μl. See Figure 5 legend for temperature program. Arrow indicates predicted band (447 nt).

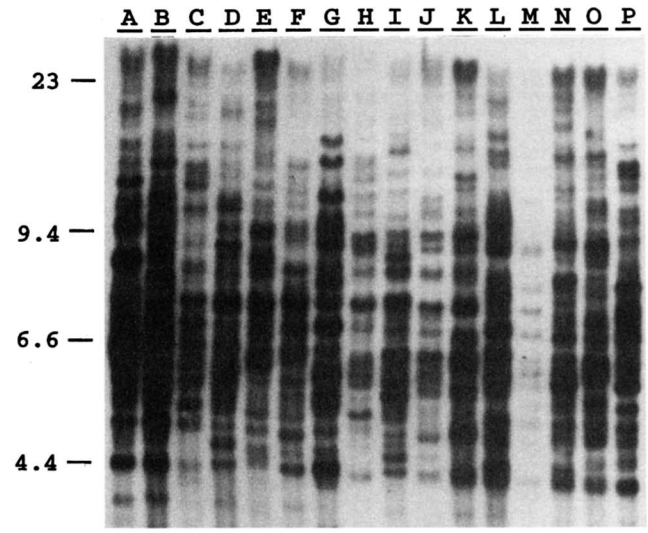

FIGURE 7.—Size diversity of *Woot* junction fragments in strains of *T. castaneum*. Genomic DNAs were digested with *Eco*RI and probed with an internal *Hin*dIII fragment (see Figure 2). Source strains and countries of origin for genomic DNAs: lane A, A^4/*Ey* (USA); B, GA-1 (USA); C, *sooty* (USA); D, NDG-2 (Canada); E, CR-1 (Costa Rica); F, *ab* (Colombia); G, Pruz-1 (USA); H, Kent (England); I, Solet (Israel); J, Ug-3 (Uganda); K, Abidjan (Cote d'Ivoire); L, Mek-1 (Peoples' Republic of China); M, Tiw-1 (India); N, Ho-Tcs (Singapore); O, Rej-1 (Philippines); P, CTC-485 (Australia).

***Woot* distribution in *T. castaneum* and other species:** Results of high-stringency Southern hybridization analysis of *Eco*RI digests of genomic DNA from various *T. castaneum* strains (Figure 7) suggests that most harbor ~25–35 copies of *Woot* per haploid genome. This conclusion is based on the expectation that the probe (labeled B in Figure 1) detects both 5′ and 3′ junction fragments, the assumption that most *Woot* insertions in these inbred strains are homozygous, and the further assumption that internally rearranged copies are rare. Figure 7 also suggests that these strains vary with respect to insertion site distribution. DNA in lane M represents the Tiw-1 strain from India. It is exceptional for its paucity of elements with high sequence similarity to *Woot*, although some weakly hybridizing bands are detected. This observation has been confirmed using DNA prepared from separate batches of M strain larvae (data not shown). The Tiw-1 strain is known to be associated with a hybrid dysgenesis-like syndrome when crossed to other *T. castaneum* strains (THOMSON *et al.* 1995). It will be interesting to assess whether *Woot* transpositional activity is associated with this phenomenon.

A PCR-based approach was utilized to assess whether other Tribolium species harbor elements closely related to *Woot*. As noted earlier, a fragment from the *pol* gene was amplified from *T. freemani* and *T. brevicornis*, and a region was sequenced that overlaps that used for the alignment in Figure 4 and extends further downstream. Sequence comparisons show that each of these Tribolium species has an element very closely related to *Woot* (Figure 8). Predicted peptides from *freemani* and *brevicornis* are 94.7 and 89.5% identical, respectively, to that of *castaneum* (Table 1). By comparison, the *Ulysses* element from *D. virilis* (the *gypsy*-class element most closely related to *Woot*) is only 30.5% identical to *Woot* over the same region (Figure 8, Table 1). We have also detected the presence of *Woot* elements in *T. confusum* by Southern hybridization and by PCR but not in *T. madens* by either method. Thus, it appears that *Woot* was present in a common ancestor of at least several existing Tribolium species, but it is not currently known whether this retrotransposon occurs outside the genus.

DISCUSSION

We have molecularly cloned and characterized a transposable element from the beetle *T. castaneum*. The organization and sequence of *Woot* indicate that it is an LTR-retroelement in the *gypsy* group. The lack of an ORF corresponding to the *envelope* gene identifies it as a retrotransposon incapable of the autonomous generation of infectious virions (see KIM *et al.* 1994; SONG *et al.* 1994; TANDA *et al.* 1994). Genomic Southerns with various strains of *T. castaneum* show that *Woot* is present in multiple copies and shows strain variability with respect to sites occupied. This conclusion is further supported by a PCR-assisted analysis of empty and occupied sites in two inbred strains (R. W. BEEMAN, unpublished results). These observations are consistent with the conclusion that *Woot* has been transpositionally active within *T. castaneum*.

At an unprecedented 3.6 kb, the *Woot* LTR is ~10 times longer than those of most other retroelements (VARMUS and BROWN 1989). The *Ulysses* element, which by *pol* sequence comparisons is most closely related to

```
Tcas   FDRVKDLFLEAVLLRYPDLNKIFYVQTDSSGYGLGAELYQIQEDGSRGVIAFASRSLKGPELNY   64
Tfre   .............H...pK...d.........r.............................   64
Tbrv   ..........T...H...qK.............................V..............   64
Ulys   IEsL.laLtt.pV.vhA.FrRp.FI.c.A.h..V..v.F.LDDEqqerp...f.ak.NkhQI..   64

Tcas   TTTEKELLGVIFALHKFRIYIQVTKIIIRTDHQALKFLSRCRLLSERLTRWTLILGQYDYEIEL  128
Tfre   ......................a.........................................  128
Tbrv   ......................a...V........R..G.............L.........  128
Ulys   Sv....c.AakL.I.R..p.VEMmpFtVi...aS.Qw.mslKd..g..A..S.e.qaFpFsMQy  128

Tcas   VKGKDNVVADILSRYPSDGEASYEYPREQPIVAMFEVNN--TPEILQLMSDLKQHQSDDPVLSQ  190
Tfre   ....................................H.--.........T.......N.....H  190
Tbrv   .......a...........D.................H.--.......Lr......lE.S..RP  190
Ulys   r..a...crH.V-.-SvE-.velt-.-.D-LLG-.QtpEFES.N.eE.IrEVmsQQGkfPdLSs  185
```

FIGURE 8.— *Woot* and *Ulysses* alignments in a region of the pol polyprotein between RNase and Integrase domains. Conceptual translations of a 573 nt (190 amino acid) segment of *Woot pol* genes from three Tribolium species were aligned with the corresponding region of *Ulysses*. The Tribolium *pol* fragments were derived by PCR using forward primer 5′-AAAATGGAGATGGGG-ACG-3′ and reverse primer 5′-ACC-CGTTGAGTTCTCGCATC-3′. Tcas, *T. castaneum* (GA-1 strain); Tfre, *T. freemani;* Tbrv, *T. brevicornis;* Dvir, the *Ulysses* element from *D. virilis* (Genbank locus DVULYSS). The *freemani* and *brevicornis* sequences are deposited in Genbank U40764 and U40765, respectively.

Woot, also has exceptionally long LTRs (2.1 kb) (SCHEINKER *et al.* 1990). However, there is no apparent similarity between the LTRs of *Woot* and *Ulysses* either in nucleotide sequence or molecular architecture. *Ulysses* LTRs show extensive repeat substructure, whereas *Woot* LTRs are almost devoid of discernable repeat motifs. In additional studies being prepared for publication, we have found that *Woot* is transcriptionally active. That is, Northern analysis indicates that the *Woot* family is actively transcribed during embryogenesis, and *in situ* hybridization to cytoplasmic transcripts shows complex temporal- and tissue-specific expression (R. W. BEEMAN, unpublished results). Internal regulatory elements responsible for this complex expression pattern could provide one explanation for the large size of the *Woot* LTRs.

The rationale for the experimental approach pursued here is that the molecular characterization of spontaneous mutations in *T. castaneum* has the potential to identify mobile transposons that might be usefully adapted for germ line transformation, cloning by transposon-tagging, and enhancer detection in this and possibly other insect species. Heretofore, no demonstrably transpositionally active element has been discovered in a non-Drosophilid insect. Examination of an apparently spontaneous mutation of the homeotic *Abdominal* gene has indeed identified a novel retrotransposon. While it is possible that this element preexisted in the mutant *Abdominal* gene as a silent insertion, it is likely that the *Woot* element is currently mobile under at least some conditions and that a new insertion caused the A^4 mutation. First, we have been unable to PCR-amplify predicted *Woot-Abdominal* junction fragments from the GA-1 stock from which A^4 arose. That is, three different primer sets that are expected to amplify junction fragments did so from the A^4 strain but gave no detectable product from the progenitor GA-1 strain (data not shown). Second, it is reasonable that the A^4 insertion would inactivate the gene. Insertion of this large 12-kb element into the transcription unit may well alter the normal processing of full-length transcripts. Moreover, a number of examples have been described in which a gene is functionally disabled because its transcripts terminate within an LTR-bearing retroelement inserted in the same orientation within the transcribed region. For the specific case of *gypsy* elements inserted within transcription units, the protein encoded by a wild-type *suppressor of Hairy-wing* (*su-Hw*) gene binds to multiple YRYTGCATAYYY sequences, resulting in frequent use of polyadenylation sites within the 5′ LTR (DORSETT 1990). Thus, when insertional mutations are also homozygous for *su-Hw* mutations, there is read-through transcription that suppresses the mutant effect. Interestingly, *Woot* has a single similar sequence, GTTGCATAC, in a position corresponding to that of the *gypsy Su-Hw* binding site (~230 nucleotides downstream of the first polyadenylation site 3′ of the putative RNA start site) (Figure 3).

TABLE 1

Similarity of *Woot* and *Ulysses* copies of a 573-nt (190-amino acid) segment of the *pol* gene

Species[a]	nt identity (%)[b]	Amino acid identity (%)[b]
Tcas	100	100
Tfre	92.8	94.7
Tbrv	82.2	89.5
Dvir	ND	30.5

[a] See Figure 8 legend for species abbreviations.
[b] All sequences are compared to Tcas and refer to those shown in Figure 8. Dvir sequence is from Genbank X56645. ND, not determined.

Two recent research breakthroughs have provoked heightened interest in insect retroelements. The first, already discussed, was the discovery that the Drosophila *gypsy* element is an infectious retrovirus (KIM *et al.* 1994). The second has been the development of a retroviral-based transformation vector with a broad host range (BURNS *et al.* 1993; LIN *et al.* 1994) including insects (J. C. BURNS, unpublished data). These and

other developments could lead to the discovery or construction of new insect retroviral pathogens or to the manufacture of new, broad spectrum gene transfer vectors. In the context of these new possibilities, further study of insect retroelements in diverse species is needed. While it is unclear whether the very large retrotransposon identified here will be technically useful, the study of recent spontaneous mutations affecting cloned genes appears to be a useful strategy for identifying mobile elements.

We thank M. S. HAAS and K. S. FRIESEN for assistance with rearing and D. M. STAUTH for assistance with DNA sequence analysis. This work was supported, in part, by U.S. Department of Agriculture, grant 92-37302-7605, and by International Human Frontier Science Program Organization Fellowship LT-568/93. The *Woot* DNA sequence is in Genbank under accession number U09586.

LITERATURE CITED

ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS and D. J. LIPMAN, 1990 Basic local alignment search tool. J. Mol. Biol. **215:** 403–410.

ALTWEGG, M., and E. KUBLI, 1980 The nucleotide sequence of histidine tRNA-gamma of *Drosophila melanogaster.* Nucleic Acids Res. **8:** 3259–3262.

ARKHIPOVA, I. R., 1995 Promoter elements in *Drosophila melanogaster* revealed by sequence analysis. Genetics **139:** 1359–1369.

ARKHIPOVA, I. R., and Y. V. ILYIN, 1991 Properties of promoter regions of *mdg1* Drosophila retrotransposon indicate that it belongs to a specific class of promoters. EMBO J. **10:** 1169–1177.

BEEMAN, R. W., J. J. STUART, M. S. HAAS and R. E. DENELL, 1989 Genetic analysis of the homeotic gene complex (HOM-C) in the beetle *Tribolium castaneum.* Dev.Biol. **133:** 196–209.

BEEMAN, R. W., J. J. STUART, R. E. DENELL, W. H. MCGAUGHEY and B. A. DOVER, 1992a Tribolium as a model insect for study of resistance mechanisms, pp. 202–208 in *Molecular Mechanisms of Insecticide Resistance,* ACS Symposium Series 505, edited by C. A. MULLIN and J. G. SCOTT. American Chemical Society Press, Washington, DC.

BEEMAN, R. W., K. S. FRIESEN and R. E. DENELL, 1992b Maternal-effect, selfish genes in flour beetles. Science **256:** 89–92.

BEEMAN, R. W., J. J. STUART, M. S. HAAS and K. S. FRIESEN, 1995 Chromosome extraction and revision of linkage group 2 in *Tribolium castaneum.* J. Hered. (in press).

BROWN, S. J., J. K. HENRY, W. C. BLACK and R. E. DENELL, 1990 Molecular genetic manipulation of the red flour beetle: genome organization and cloning of a ribosomal protein gene. Insect Biochem. **20:** 185–193.

BURNS, J. C., T. FRIEDMANN, W. DRIEVER, M. BURRASCANO and J. K. YEE, 1993 Vesicular stomatitis virus G glycoprotein pseudotyped retroviral vectors: concentration to very high titer and efficient gene transfer into mammalian and nonmammalian cells. Proc. Natl. Acad. Sci. USA **90:** 8033–8037.

DOOLITTLE, R. F., D. F. FENG, M. S. JOHNSON and M. A. MCCLURE, 1989 Origins and evolutionary relationships of retroviruses. Quart. Rev. Biol. **64:** 1–30.

DORSETT, D., 1990 Potentiation of a polyadenylation site by a downstream protein-DNA interaction. Proc. Natl. Acad. Sci. USA **87:** 4373–4377.

ECHALIER, G., 1989 Drosophila retrotransposons: interactions with genome. Adv. Virus Res. **36:** 33–105.

FRIESEN, P. D., and M. S. NISSEN, 1990 Gene organization and transcription of *TED,* a lepidopteran retrotransposon integrated within the baculovirus genome. Mol. Cell. Biol. **10:** 3067–3077.

HALISCAK, J. P., and R. W. BEEMAN, 1983 Status of malathion resistance in five genera of beetles infesting farm-stored corn, wheat and oats in the United States. J. Econ. Entomol. **76:** 717–722.

HATFIELD, D. L., J. G. LEVIN, A. REIN and S. OROSZLAN, 1992 Translational suppression in retroviral gene expression. Adv. Virus Res. **41:** 193–239.

HENIKOFF, S., 1984 Unidirectional digestion with exonuclease III creates targeted deletion breakpoints for DNA sequencing. Gene **28:** 351–359.

KIM, A., C. TERZIAN, P. SANTAMARIA, A. PELISSON, N. PRUD'HOMME *et al.,* 1994 Retroviruses in invertebrates: the *gypsy* retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster.* Proc. Natl. Acad. Sci. USA **91:** 1285–1289.

LIN, S., N. GAIANO, P. CULP, J. C. BURNS, T. FRIEDMANN *et al.,* 1994 Integration and germ-line transmission of a pseudotyped retroviral vector in zebrafish. Science **265:** 666–669.

MARLOR, R. L., S. M. PARKHURST and V. G. CORCES, 1986 The *Drosophila melanogaster gypsy* transposable element encodes putative gene products homologous to retroviral proteins. Mol. Cell. Biol. **6:** 1129–1134.

NAKAKITA, H., O. OMURA and R. G. WINKS, 1981 Hybridization between *Tribolium freemani* (Hinton) and *Tribolium castaneum* (Herbst) and some preliminary studies on the biology of *Tribolium freemani* (Coleoptera: Tenebrionidae). Appl. Entomol. Zool. **16:** 209–215.

O'BROCHTA, D. A., W. D. WARREN, K. J. SAVILLE and P. W. ATKINSON, 1994 Interplasmid transposition of Drosophila *hobo* elements in non-drosophilid insects. Mol. Gen. Genet. **244:** 9–14.

PELISSON, A., S. U. SONG, N. PRUD'HOMME, P. A. SMITH, A. BUCHETON *et al.,* 1994 *Gypsy* transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the Drosophila *flamenco* gene. EMBO J. **13:** 4401–4411.

ROBERTSON, H. M., 1993 The *mariner* transposable element is widespread in insects. Nature **362:** 241–245.

SANGER, F., S. NICKLEN and A. R. COULSON, 1977 DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:** 5463–5467.

SCHEINKER, V. S., E. R. LOZOVSKAYA, J. G. BISHOP, V. G. CORCES and M. B. EVGENEV, 1990 A long terminal repeat-containing retrotransposon is mobilized during hybrid dysgenesis in *Drosophila virilis.* Proc. Natl. Acad. Sci. USA **87:** 9615–9619.

SONG, S. U., T. GERASIMOVA, M. KURKULOS, J. D. BOEKE and V. G. CORCES, 1994 An env-like protein encoded by a Drosophila retroelement: evidence that *gypsy* is an infectious retrovirus. Genes Dev. **8:** 2046–2057.

STUART, J. J., S. J. BROWN, R. W. BEEMAN and R. E. DENELL, 1993 The Tribolium homeotic gene, *Abdominal,* is homologous to *abdominal-A* of the Drosophila Bithorax Complex. Development **117:** 233–243.

TANDA, S., J. L. MULLOR and V. G. CORCES, 1994 The Drosophila *tom* retrotransposon encodes an envelope protein. Mol. Cell. Biol. **14:** 5392–5401.

THOMSON, M. S., K. S. FRIESEN, R. E. DENELL and R. W. BEEMAN, 1995 A hybrid incompatibility factor in *Tribolium castaneum.* J. Hered. **86:** 6–11.

VARMUS, H., and P. BROWN, 1989 Retroviruses, pp. 53–108 in *Mobile DNA,* edited by D. E. BERG and M. M. HOWE. American Society of Microbiology, Washington, DC.

XIONG, Y., and T. H. EICKBUSH, 1990 Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. **9:** 3353–3362.

YUKI, S., S. INOUYE, S. ISHIMARU and K. SAIGO, 1986 Nucleotide sequence characterization of a Drosophila retrotransposon, *412.* Eur. J. Biochem. **158:** 403–410.

Communicating editor: M. J. SIMMONS